Word/phrase filtering with Policy Patrol

Policy Patrol can search emails for words and phrases by making use of word/phrase filters. For anti-spam purposes, the program includes a Word Block List and Word Allow List. In order to increase accuracy and avoid false positives, Policy Patrol allows you to specify word score, case sensitivity and multiple count for each word. In addition Policy Patrol offers word pattern matching through the use of regular expressions.

Configuring Word/Phrase Filters

Follow the next steps to create a Word/Phrase filter:

1. Go to **Settings** > **Filters** > **<folder>** and click **New...**.

2. When asked which type of filter you wish to create, select **Word/phrase Filter**. Click **Next**. Enter the word(s) or phrases to be included in the filter (see paragraph 'Configuring words/phrases'). When you are ready adding words, click **Next**.

3. Enter a name for the filter and any additional comments. When you are done, click **Finish** to create the filter.

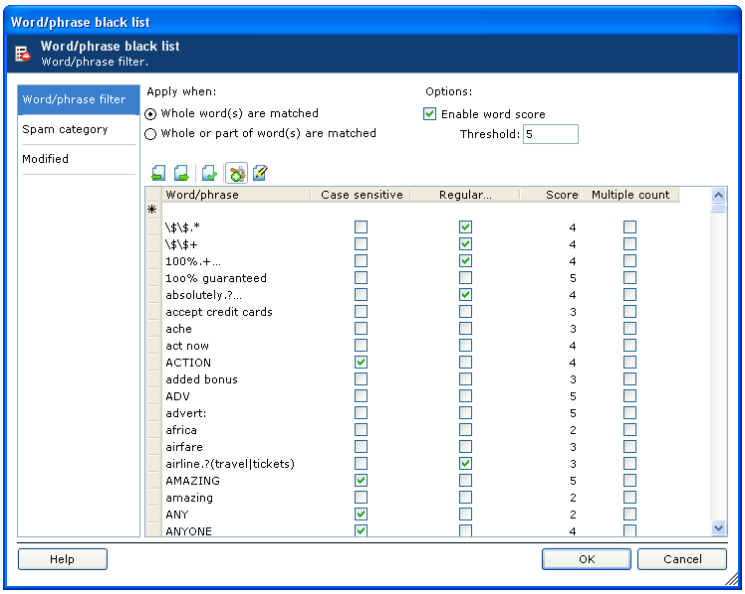
Word Block lists and Allow Lists

Word Block lists and Word Allow Lists are preconfigured filters used for spam blocking. To edit the words in these filters, go to **Email Threat Protection** > **Anti-spam** > **Block** > **Word Block List** and **Anti-spam** > **Allow** > **Word Allow List**. You will be able to edit the words in the lists (see paragraph 'Configuring words/phrases'). When you are done, click **OK** to save the changes.

Note: Use caution when editing Word Block Lists and Word Allow Lists. Words that are too broad in the block lists can result in false positives, whereas unspecific words in the allow list can lead to false negatives. If you enter your company name in the allow list and your company name is the same as your domain, e.g. Microsoft and microsoft.com, make sure that the option **Whole words are matched** is selected in the white list. Many spammers include your email address in the subject or body of the email. If you select this option, Policy Patrol will only allow emails that include your company name as a separate word, not as part of a domain.

Configuring words/phrases

For each word entry you can apply a word score and select whether it should be case sensitive and/or counted multiple times.

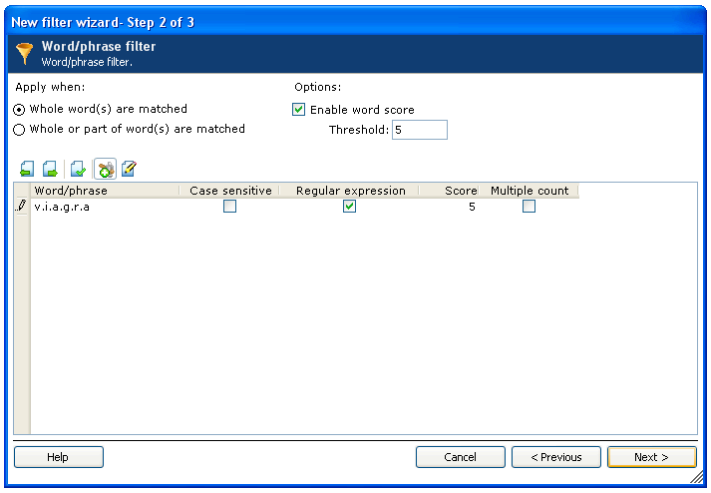


Case sensitivity

If you check the **Case sensitive** option, this means that Policy Patrol will only check for the word in the same case. This can be useful for certain spam or chain letters for instance, that tend to use a lot of capitals. For instance if a mail includes CLICK HERE in capitals there will be a good chance that the mail is spam. However, click here in lower case might be more innocent. By using the case sensitive option in combination with the word score option you could add both variations, applying a higher score to the upper case version. Remember though that if you enter a case sensitive and non-case sensitive version of the same word, and the word in the email matches the case sensitive version, the word will be counted twice since it will match both the case sensitive and non-case sensitive entry in the filter.

Regular expression

If the entry is a regular expression tick the box **Regular expression**. Regular expressions allow you to match a word pattern instead of an exact word. This means that by making use of regular expressions you can stop spammers trying to circumvent content filters by adding characters within words, such as v*i*a*g*r*a or c-l-i-c-k h-e-r-e. Furthermore you can detect word variations such as r@tes and l0ans. Policy Patrol includes an extensive spam words filter that makes use of many regular expressions to detect variations of spam words.



Below are a couple of expressions that can be used:

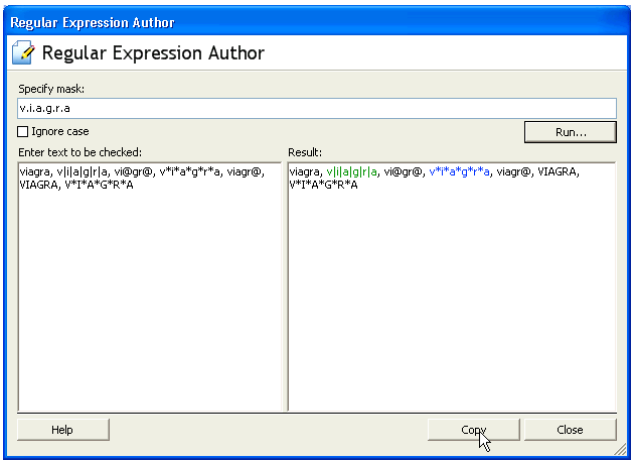
Regular expression	Meaning
\b	Word boundary
.	Any character
*	Previous character 0 or more times
+	Previous character 1 or more times
[a,b]	Character a or b

For instance, if you enter `vi[a|@]gr[a|@]` this will find the word viagra, vi@gra, viagr@ and vi@gr@. If you enter `v.i.a.g.r.a`, this will find the words v*i*a*g*r*a, v|i|a|g|r|a and v-i-a-g-r-a. Note that the options **Whole word(s) are matched** and **Whole or part of word(s) are matched** do not apply to regular expressions since this can be indicated in the regular expression itself. The case sensitivity, word score and multiple count options do apply.

Note: Be cautious when using the * sign in word entries. If the word is not marked as a regular expression, the * is seen as a wildcard for any character. This means that if you enter the word v*i*a*g*r*a this will not only find v/i/a/g/r/a and v-i-a-g-r-a, but also the phrase: Victor is a great person. If you enter the word v*i*a*g*r*a and check the regular expression tick box, this means that the entry will trigger on all words since the * sign means 0 or more of the previous character.

Policy Patrol includes a Regular Expression Author to help you create and test your regular expressions. Follow the next steps to use the Regular Expression Author:

1. Click on the **Regular Expression Author** icon in the toolbar .



2. In **Specify mask**, enter your regular expression, for instance `v.i.a.g.r.a`. If you wish to ignore case, select the option **Ignore case**.

3. In the left dialog, enter the sample text to be checked for the regular expression.

4. Click on **Run**. The words that match the regular expression will be colored green and blue alternately. For instance, in the example above, you can see that the regular expression v.i.a.g.r.a matches v*i*a*g*r*a, but not viagra or vi@gra.

5. If the result is not as you had intended, alter the regular expression and press **Run** again. If your regular expression produced the intended results, press **Copy** and **Close**. Now paste the regular expression into the word/phrase filter and tick the box **Regular expression**.

Word score

If you wish to use word score you must check **Enable word score**. For each word you will now be able to apply a word score. In the **Threshold** dialog box, specify the word score threshold that must be met to trigger the action. You can also apply a negative word score. If the total score for the words found in the message equals or exceeds the word score threshold, the action or policy will trigger. In other words, if you enter two words in the filter with both a score of 5, and the threshold is set to 10, the policy or action will trigger when at least both entries are found in the email or, if multiple count is ticked, two instances are found of one of the entries. If you do not wish to use word scores in the filter, uncheck **Enable word score**. If you do not enable word score, messages that include one or more of the words will trigger the action.

Negative word score

You can also apply a negative word score. This can be useful to eliminate some words that can be used innocently. For instance you might assign the word 'breast' a word score of 5, and assign the words 'baby' or 'chicken' a minus 5 score. You can also add a negative score to words that indicate legitimate emails such as your company name and your product or service name. By setting different word scores and applying negative scores for certain words, it is possible to closely identify the content of emails and in doing so greatly decrease the occurrence of false positives (i.e. wrongly triggered policies).

Multiple count
If you wish every instance of the word to be counted, check the box **Multiple count**. For example, if this box is enabled and you receive an email message that contains the word 'debt' three times, and you applied word score of 5 to this word, the total word score would be 15. If you did not check this box, the word will only be counted once and the total score would be 5.

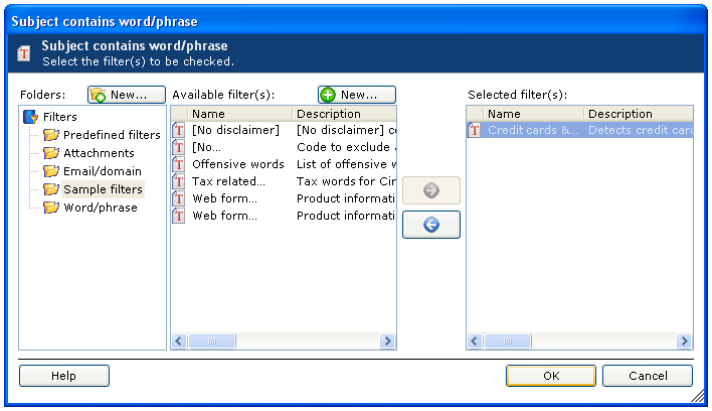
Import/Export
You can import lists from .txt files by clicking on **Import**, browsing to the appropriate file and clicking **Open**. The format should be as follows: Word[TAB]Case sensitive[TAB]Regular expression[TAB]Score[TAB]Multiple count. The word/phrase and score values must be entered. For the other options, either 1 (enabled) or 0 (disabled) must be entered. For instance, if you wish to add the case sensitive word CLICK HERE with a word score of 5 and multiple count, you must enter it in the text file as follows: `CLICK HERE  1  0  5  1`. For every word or phrase you need to start a new line. To export the words in the filter, click **Export**, enter a file name and select **OK**.

Whole or part of words
Select whether to apply the filter when **Whole word(s) are matched** or when **Whole or part of word(s) are matched**. The first option allows you to specify more precisely which words must trigger a policy. For instance, if you select that **Whole or part of word(s) are matched** and you enter the word 'sex' in the filter, this will also include the words 'Sussex' and 'sextant'. If you select **Whole word(s) are matched**, the policy will trigger on the word 'sex' but not on 'Middlesex'.

Configuring word/phrase filtering in the policy

Follow the next steps to configure a policy that checks for words or phrases:

1. Go to **Email Content Policies** > **<folder>** and click **New**.
2. Select the users for the policy and click **Next**.
3. Select which messages you want Policy Patrol to check and click **Next**.
4. Select **Trigger policy if following conditions are met**. For word/phrase filtering you can select one or more of the following conditions:
   - **Subject contains word/phrase** (searches the subject of the email)
   - **Body contains word/phrase** (searches the body of the email)
   - **Attachment contains word/phrase** (searches the attachment of the email)
5. After selecting one of the above options, click on the word/phrase link in the description and select the filter(s) you want to check.

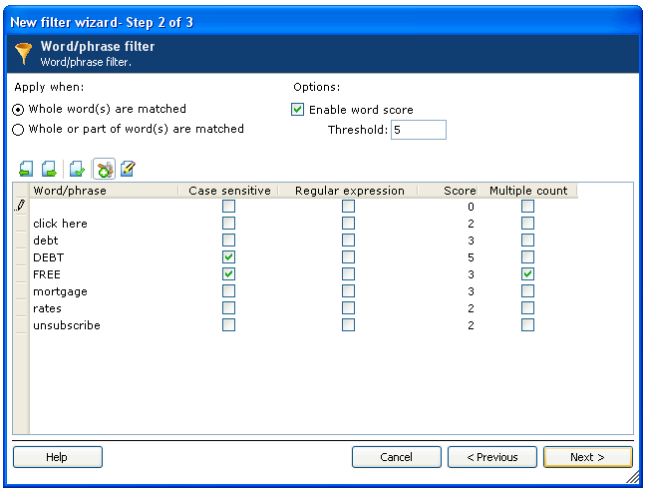


If you selected to search the body and you wish to search for HTML tags, check the option **Check HTML tags**. This can be useful if you want to check for scripts by searching for the `<SCRIPT>` tag. However, if you wish to check normal text, you must not select this option since this will produce unwanted results. Click OK and **Next**.

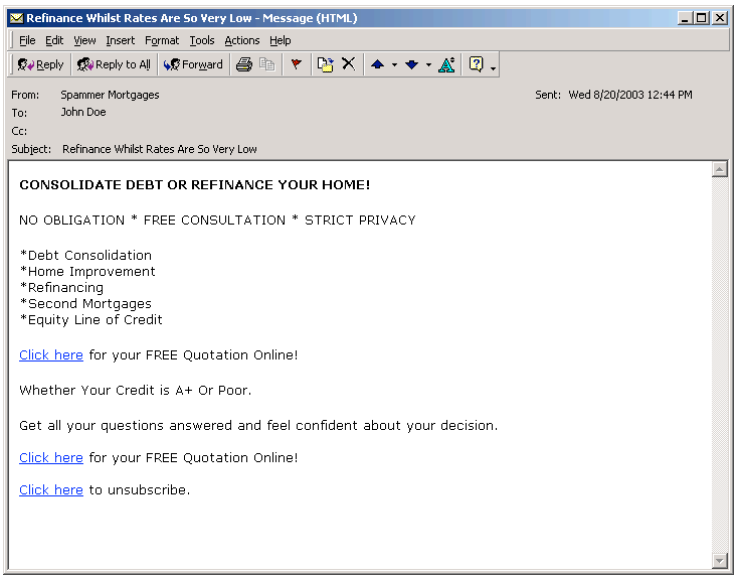
6. Specify any exceptions and click **Next**.
7. Configure the Actions and click **Next**.
8. Configure Scheduling if you wish and click **Next**.
9. Enter a name and description for the policy and click **Finish**.

Example

In order to explain the word score functionality, an example is discussed here. In the example you have configured your word/phrase black list as shown in the screen below (note that we recommend using the sample black list, this is just used as an example). You have selected that whole word(s) should be matched and you have enabled word score with a threshold of 5. This means that the message is considered spam when the total score of words found in the body or subject equals 5 or more.



A spam message is sent to your organization as shown in the screen below.



Policy Patrol quarantines the message. You now go to **Monitoring folders** > **<folder name>** and select the quarantined message. In the bottom pane, click on the button *Policies report*. The Policies report will show that the policy triggered and will list the words found and their scores. In the report you can see that Policy Patrol found words in the subject (score of 2) and body of the message (score of 21). Since you configured a word score threshold of 5, the message was considered as spam.

Test Word filter rule	Yes
Word/phrase from filter Test filter	*21*
click here	*2*
debt	*3*
DEBT	*5*
FREE	*9*
unsubscribe	*2*

We now take a closer look at how Policy Patrol determined the word score, by examining the email message (the message is displayed below with the words from the filter in bold).

⇒ The non case sensitive word `rates` was found in the subject. (score=2)

⇒ Policy Patrol finds three instances of the case sensitive word `FREE`. Since multiple count is selected for this word, the word score is counted 3 times. (score=9)

⇒ One instance of the non case sensitive word `unsubscribe` is found. (score=2)

⇒ Policy Patrol finds two instances of the non case sensitive word `debt`. However since multiple count is not enabled for this word, it is only counted once. (score=3)

⇒ One instance of the case sensitive word `DEBT` is found (Note that if the same word matches multiple word/phrase entries in the filter, each entry will be counted). (score=5)

⇒ Three instances of the non case sensitive word `click here` are found in the email body. However since multiple count is not enabled, the word is only counted once. (score=2)

⇒ Policy Patrol does not find the word `mortgage` in the email, since you checked the option **Whole word(s) are matched** in the Spam words filter, and the email message contains the word `Mortgages`. If you had selected **Whole or part of word(s) are matched**, Policy Patrol would have counted this word as well. This option must be used with caution though, since it can result in some false positives. For instance, if you enter the word `sex` in a Word/Phrase filter and you enable the option **Whole or part of word(s) are matched**, Policy Patrol will also find the word `sex` in words such as Sussex, Middlesex and sextant.

Subject:	Refinance Whilst **Rates** Are So Very Low
Body:	CONSOLIDATE **DEBT** OR REFINANCE YOUR HOME! NO OBLIGATION * **FREE** CONSULTATION * STRICT PRIVACY ***Debt** Consolidation *Home Improvement *Refinancing *Second Mortgages *Equity Line of Credit **Click here** for your **FREE** Quotation Online! Whether Your Credit is A+ Or Poor Get all your questions answered and feel confident about your decision. **Click here** for your **FREE** Quotation Online!

	Click here to **unsubscribe**.

Troubleshooting

Q: My policy that searches for words/phrases is triggering erratically

A: Check whether the option **Check HTML tags** is selected in Policy Properties > Policy conditions tab > word/phrase filter link. If this option is selected Policy Patrol will search HTML tags as well as text and might produce unwanted results if used for checking normal text. For instance, spammers frequently use comment tags within the text (which are not displayed on screen) to circumvent content filters. By default, Policy Patrol will ignore these tags, however if **Check HTML tags** is selected it will include these tags in the search.

More information

⇒ For more information on how to use regular expressions, please download the document 'Using regular expressions in Policy Patrol' from: http://www.policypatrol.com/assets/pp10-regularexpressions.pdf.

⇒ For more information on how to configure Policy Patrol, please download the appropriate quick start guide and manuals from: http://www.policypatrol.com/documentation.

⇒ If you still have questions after reading this document, please consult our online knowledge base at http://support.redearthsoftware.com, or send an email to support@redearthsoftware.com.

Contacting Red Earth Software

Red Earth Software, Inc.
4845 Pearl East Circle, Ste 101
Boulder, CO 80301
United States
Toll-free: 1-800-921-8215
Phone: (720) 377-3728
Fax: (720) 554-7950
Sales: sales@redearthsoftware.com
Support: support@redearthsoftware.com

Red Earth Software (UK) Ltd
20 Market Place
Kingston-upon-Thames
Surrey KT1 1JP
United Kingdom
Tel: +44-(0)20-8328 9830
Fax: +44-(0)20-8711 5771
Sales: sales@redearthsoftware.co.uk
Support: support@redearthsoftware.co.uk

Policy Patrol® is a registered trademark of Red Earth Software®. Copyright © 2001- 2014 by Red Earth Software.